

**Dirty Hands on Troubled Waters:
Sanitation, Access to Water and Child Health in Ethiopia**

Wondimu S. Manalew^a | Vidhura S. Tennekoon^a

^a Department of Economics, Indiana University Purdue University Indianapolis (IUPUI), School of Liberal Arts, 425 University Boulevard, Cavanaugh Hall, Indianapolis, IN 46202, USA.

Correspondence

Wondimu S. Manalew, Department of Economics, Indiana University Purdue University Indianapolis (IUPUI), School of Liberal Arts, 425 University Boulevard, Cavanaugh Hall, Indianapolis, IN 46202, USA. Email: wmanalew@iu.edu.

Abstract

In this paper, we investigate the impact of access to drinking water sources and sanitation facilities on the incidence of diarrheal diseases among children below five in Ethiopia using the propensity score matching technique with a polychotomous treatment variable. We find that among the water sources traditionally considered as improved, only water piped into dwelling, yard or plot leads to a large percentage point reduction in diarrhea incidence. The other water sources, generally believed as clean, are not effective in reducing diarrhea even compared to some of the unimproved water sources. We also find that some unimproved water sources and sanitation facilities are less inferior than they are believed to be. These results suggest that the traditional way of categorizing different types of improved and unimproved water sources and sanitation facilities into a dichotomous variable, “improved” or “unimproved”, could be misleading as it masks the heterogeneous effects of the water sources and the sanitation facilities.

JEL Codes: I10; I15; O10; O20; O55.

Keywords: access to water; sanitation; child health; diarrhea; Ethiopia; propensity score matching.

This is the author's manuscript of the article published in final edited form as:

Manalew, W. S., & Tennekoon, V. S. (2019). Dirty hands on troubled waters: Sanitation, access to water and child health in Ethiopia. *Review of Development Economics*, 23(4), 1800–1817. <https://doi.org/10.1111/rode.12604>

1. Introduction

The World Health Organization (WHO) estimates that more than half million children under the age of five die each year from diarrhea, mainly in developing countries.¹ Unsafe drinking water and unimproved sanitation facilities are widely thought to be the major causes of diarrheal diseases. According to Centers for Disease Control and Prevention (2016), an estimated 88 percent of diarrheal deaths worldwide are attributable to unsafe water supply, inadequate sanitation and insufficient hygiene conditions.

The strong connection between diarrheal diseases and access to clean water and sanitation facilities has motivated public programs to expand access to clean water and improved sanitation especially in developing countries where initial access to these facilities is very low (Jalan and Ravallion, 2003). Some developing countries have progressed more than others in increasing access. Ethiopia, for example, achieved the largest decrease in the proportion of the population practicing open defecation (from 92 per cent in 1990 to 29 per cent in 2015), a reduction over five times greater than the sub-Saharan Africa average for the same period (UNICEF and WHO, 2015). The percentage of people using improved water sources also has increased during the same period. Despite the remarkable progress in increasing access, only 28 percent of the nearly 100 million Ethiopian population had access to an improved sanitation facility in 2015 while a slightly larger share of the population defecated openly. The remaining 43 percent of the population either used a shared facility or an unimproved facility². Access to improved water sources is also low. In 2015, only 57 percent of the population had access to an improved water source (UNICEF and WHO, 2015).

¹ <http://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease>

² This is a remarkable progress compared to the situation in 2011, when our study data sample was collected. As our data shows 51.5 percent of the Ethiopian households defecated openly. Only 8.2 percent had an improved facility while the remaining households used either an unimproved or a shared facility.

In this paper, we investigate the impact of access to various drinking water sources and sanitation facilities on incidence of diarrhea among children aged below five years in Ethiopia. Similar studies have been conducted in other countries including India (Jalan and Ravallion, 2003; Fan and Mahal, 2011; Kumar and Vollmer, 2012), Bangladesh (Begum et al., 2011), Senegal (Novak, 2014), Philippines (Capuno et al., 2015), and Egypt (Roushdi and Sieverding, 2016). While most of these studies focus on the impact of water sources, Begum et al. (2011) and Capuno et al. (2015), like this paper, investigate the impact of both water sources and sanitation facilities on incidence of diarrhea.

Except for the three studies on India, which use data from various household surveys, the other studies use the Demographic and Health Survey (DHS) data for their analyses. However, ignoring detailed information DHS provides on different types of water sources and sanitation facilities, these studies, except Novak (2014), classify water sources and sanitation facilities under two broad groups as “improved” or “unimproved”³. Novak (2014) argues that using this broad classification restricts different water sources to have the same estimated effect on child health. Novak’s argument is reasonable given that the improved sources include presumably treated commercial supplies of water piped into dwelling, yard or the plot, ground water collected from protected wells, protected springs and tube wells as well as rain water. Each of these water sources is likely to be contaminated through very different channels and therefore the effect of those different sources on the incidence of diarrhea need not be similar.

To correct for such an aggregation error that could mask the possible heterogeneous effects of the different water sources, Novak classifies the improved water sources into four categories: piped into dwelling, piped into yard or plot, public tap and protected well. Nevertheless, she groups

³ See Table 1 and Table 2 for details on different types of waters sources and sanitation facilities.

all the unimproved water sources in a single category regardless of the type of the water source and the location. According to Novak (2014, pp. 434), aggregating the unimproved water sources is reasonable because,

*“(1) organizations and governments do not typically build unimproved water sources which makes estimating their heterogeneous impacts irrelevant, and
(2) unimproved water sources will have a similar treatment effect.”*

In this study, we argue that grouping all the unimproved sources together too could mask the heterogeneous effects of upgrading from different unimproved water sources to improved water sources. This is because, similar to improved sources, unimproved sources also include sources of ground water such as unprotected wells and unprotected springs, sources of surface water (rivers, dams, lakes, ponds, streams, canals, and irrigation channels) as well as commercial supplies of water transported using donkey carts, motorized vehicles and tanker trucks. The channels through which each of these different sources of unimproved water could be contaminated are not the same. Therefore, they are likely to differ in terms of their effect on the incidence of diarrhea.

Although we agree with Novak that governments and organizations do not typically build unimproved sources, we argue that they would want to know replacing of which unimproved sources would be more cost effective and would result in higher reduction in diarrheal prevalence. For example, replacing ‘unprotected well’ with a particular type of improved water source may differ in its cost-effectiveness from replacing ‘river water’.

Another issue with the categorization of several heterogeneous water sources under two broad categories is that the classification of some sources as improved or unimproved itself could be questionable at least in some country settings. For example, the difference between a protected

well and an unprotected well or between a protected spring and an unprotected spring could not only be subtle but also is likely to be subjective. Therefore, two survey respondents may classify the same well or spring differently based on their subjective evaluations making the two groups more similar than they should be and thereby masking the effect of better water source on preventing diarrhea.

In this paper, we classify both the improved and unimproved water sources into different categories. Accordingly, our analysis disaggregates water sources into six categories; three improved categories and three unimproved categories. We also decompose household sanitation facilities into three categories. Our results suggest that the disaggregation of the water sources and the sanitation facilities provides policy relevant information that would otherwise have not been known.

Finally, a policymaker who intends to reduce the incidence of diarrhea in Ethiopia under budget constraints may want to identify her priorities as to whether providing a toilet is more cost effective than providing access to clean water or vice versa. The policymaker may also want to compare the effectiveness of providing both an improved water source and an improved sanitation facility to the same household versus providing each facility to two different households. To test this, we investigate the effect of an improved water source on those with and without a toilet facility as well as the effect of an improved sanitation facility with or without an improved water source and find mixed evidence.

The rest of the paper is structured as follows. In section 2 we present our model and discuss the methodology employed. Section 3 provides a brief description of the data used for the analysis. In section 4 we present and discuss the estimation results. Section 5 summarizes and concludes.

2. The Model

2.1. Notation and Background

As first formalized in Rubin (1974), the estimation of causal effects, whether from data in a randomized experiment or an observational study, is inherently a comparison of potential outcomes on individual units, where a unit is a physical object (e.g. a person or an institute) at a specific point in time (Stuart and Rubin, 2004). The causal effect for unit i is the comparison of unit i 's outcome if it receives the treatment (unit i 's potential outcome under treatment), $Y_i(1)$, and unit i 's outcome under the control conditions (unit i 's potential outcome under control), $Y_i(0)$. The “fundamental problem of causal inference” (Rubin, 1978) is that, for each unit, we can observe only one of these (two) potential outcomes, because each unit will either receive the treatment or face control conditions, not both. If both potential outcomes were observable, the causal effect of the treatment for the i^{th} observation would be the difference $Y_i(1) - Y_i(0)$ and the average treatment effect (ATE), the average of the difference between the observed and unobserved potential outcomes for each subject, would be the expected value of this difference, as given in equation (1) below.

$$ATE = E[Y_i(1) - Y_i(0)] \quad (1)$$

As we have already pointed out, we only observe either $Y_i(1)$ or $Y_i(0)$ for any single observational unit i . Given this fact, obtaining causally interpretable treatment effects, requires predicting the other unobserved potential outcome, which enables the comparison of the potential outcomes under treatment and control conditions. The main obstacle faced by researchers in social sciences that aim to estimate the effect of an intervention (treatment) is eliminating the selection bias, where the units which receive the treatment may have different characteristics from those in the control condition (Stuart and Rubin, 2004). To eliminate/minimize the selection bias and

identify the ATE correctly or with minimum error, pre-existing differences between the groups must be controlled using appropriate econometric techniques.

Various econometric methods have been developed to deal with the selection bias, matching being one of those techniques. This technique is used frequently to reduce or avoid biases that result from the absence of ideal counterfactuals. The matching technique pairs treated individuals with untreated individuals based on observable characteristics allowing comparisons with proper counterfactuals and thereby avoiding or minimizing selection bias. Meaningful comparisons between the outcomes of the two groups, however, require some assumptions. The first assumption is the conditional independence (CI) assumption. This assumption states that the outcome is independent of the treatment assignment mechanism, conditional on pre-treatment observable characteristics. This assumption can be formally expressed as follows:

$$Y_i(0), Y_i(1) \perp D_i | X_i = x_i, \forall x \in \chi \quad (2)$$

where $Y_i(0)$ and $Y_i(1)$ are as defined above and D_i is the treatment status such that $D_i = 1$ if the individual is treated and 0 otherwise, and X_i is a set of observable pre-treatment characteristics in characteristic space χ .

The conditional independence assumption implies that once we control for the effects of regressors X_i , treatment and outcomes are independent. The X_i here should include all the variables that determine both the treatment and the outcome. Any other observable or unobservable factors not included in X_i may determine the outcome if they are independent of the treatment and may determine the treatment if they are independent of the outcome. If the CI assumption holds, matching untreated individuals to treated individuals that have the same pre-treatment observable characteristics ensures that the selection bias is eliminated. On that basis, the ATE can be measured accurately using equation (1).

In this paper, since we deal with more than two different types of water sources and sanitation facilities, identifying the causal effect requires observing the health status of the child in multiple states. We need the CI in this case as well. Following Lechner (2002) and Hirano and Imbens (2004), the CI assumption in the binary framework of matching can be extended to the case of multiple (J) treatment states.

For multiple-treatments case, the CI assumption is given as

$$Y_i(0), Y_i(1) \perp T_i | X_i = x_i, \forall x \in \chi \quad (3)$$

where T_i is the treatments status with $T_i = \{0, 1, 2, \dots, J\}$.

The other requirement necessary to employ the matching technique is the overlap of the distributions of covariates in control and treatment groups. Without sufficient overlap of two groups (common support), it is hard to estimate the average treatment effects precisely.

2.2. Propensity Scores and the Estimation of Average Treatment Effects

In this paper, we use the most commonly used matching method, propensity score matching (Rosenbaum and Rubin, 1983). Our estimation of the treatment effects using this method involves two steps. First, we use the multinomial logistic regression to predict the probabilities (propensity scores) of using each water source and sanitation facility for each treatment relative to a base category. After obtaining the propensity scores for the entire sample, we estimate pairwise propensity scores following Lechner (2002) and using (4) below.⁴

$$\hat{P}_{t_i|t_i, u_i}(x_i) = \frac{\hat{P}_{t_i}(x_i)}{\hat{P}_{t_i}(x_i) + \hat{P}_{u_i}(x_i)} \quad (4)$$

where t_i and u_i represent “treated” and “untreated” water sources, respectively.⁵

⁴ We are interested in the pairwise propensity scores because we want to estimate the effect of moving from each water source to another.

⁵ Propensity scores, estimating using a multinomial logit model and equation (4) above, however, are qualitatively not different from those estimated using separate univariate logit models of each combination of water sources. This is because of independence of irrelevant alternatives (IIA) property of the logit model.

In the second stage, we match each observation in the treatment group with an observation in the control group and each observation in the control group with one in the treatment group using the nearest neighbor matching algorithm with replacement in order to estimate the ATEs of different pairs of water sources. The treatment here is defined as replacing one source of drinking water with another, replacing an unimproved source with an improved source when at least one of the two sources belong to unimproved category. When pairs of two improved sources are compared, the choice of the treatment and control outcomes, and therefore the definition of the outcome is arbitrary. This is because the better source of the two is not known in advance⁶.

Propensity score matching, compared to direct matching, helps to overcome the ‘curse of dimensionality’ which is a term used to explain the technical difficulties and other related issues involved when each observation in the treatment group has to be matched with an identical counterfactual from the control group (or vice versa). First, this is computationally highly taxing and consumes a lot of processing power and time as the number of covariates and the number of observations increase. This is because every observation in the opposite group has to be inspected before finding a match and this process has to be repeated as many times as the number of covariates in the model. Second, this often causes the sample size to drop significantly after matching because an exact match is hard to find so that the values of many covariates are either equal or close enough. The propensity score, which is constructed as a composite index which includes information from all relevant covariates solves this problem and maps the multidimensional matching procedure to a one-dimensional procedure. Since the propensity scores are composite indicators representing all covariates to be controlled for, the CI assumption changes as follows.

⁶ As a result, same household using a particular water source could be in a control or a treatment group depending upon the pairwise comparison of interest.

$$Y_i(0), Y_i(1) \perp T_i | \hat{P}_i(x_i), \forall x \in \chi \quad (5)$$

The nearest neighbor matching technique that we use to produce our baseline results (Tables 4-6) does not restrict two very different propensities being matched if they are the closest possible matches. Therefore, as a robustness check (Tables 7-9), we set a caliper (maximum distance between propensity scores acceptable for matching) using the formula,

$$c = \alpha \sqrt{(\sigma_t^2 + \sigma_u^2)/2} \quad (6)$$

where c is the caliper size, σ_t^2 and σ_u^2 are the variances of propensities of the treated and the untreated observations, respectively, and $\alpha = 0.2$ is a fixed parameter to be chosen by the researcher which corresponds to a tight matching that removes “practically all the bias” (Cochran and Rubin, 1973)⁷. In each case, only the observations within the range of common support were matched with the other group⁸.

3. Data

We use data from the Demographic and Health Survey (DHS) conducted in Ethiopia in 2011⁹. Field work of this survey was conducted during the six months period from December 2010 to May 2011 and a nationally representative sample of 17,817 households were selected for the survey. Of these, 16,702 households successfully completed the interview. Focus of our study is 8,980 children under 5 years of age lived in those households¹⁰. DHS assigned a sampling weight to each child in the sample because the implementation of this survey was nonrandom, and the

⁷ In their examples, choosing $\alpha = 0.2$ eliminates 98-99% of the bias.

⁸ For implementing this, we considered the larger of the minimum values of estimated propensity scores in treatment and control groups as the lower bound of common support and the smaller of the maximum values as the upper bound.

⁹ This was the most recent DHS dataset available when this research study was started.

¹⁰ The DHS contained data collected about 11,654 children below 5. However, 2,674 of these observations lacked required information (water source and the type of sanitation facility used, for example) and were not used in our analyses. The observable characteristics of these children, however, were not different from those included in our analysis.

probability of being sampled was different for each child. We use these weights when estimating the first stage logit models.

Outcome Variable

The outcome variable of interest was the incidence of diarrhea among our target group during the two weeks prior to the interview date. Overall, this rate is 15.1 percent. However, the rate of incidence varies with type of water source used for drinking (Table 1) and the availability of sanitation facilities (Table 2). Diarrhea incidence also varies with area of residence, age of the child, parents' level of education, household wealth, among other things.

Treatment variables

Respondents in the survey were asked to select their main drinking water source from a list of 13 different water sources which are based on the definitions of WHO/UNICEF (2015). First, we categorize these 13 water sources into two groups, 'improved' and 'unimproved', as done traditionally. In this case, our first stage outcome variable ('water source') is a binary variable that takes 1 if the water source used by the household is 'improved' and 0 otherwise. Second, we divide the improved water sources into three subcategories ('piped into dwelling/ yard/plot', 'public tap/standpipe' and 'other improved') leaving all the unimproved water sources in one category as 'unimproved'. This approach is similar to Novak (2014) except that water piped into dwelling and water piped in to the yard/plot are merged in our case due to data limitations. The 'other improved' category includes protected wells, tube wells and bore holes, protected springs and rain water. Finally, we divide not only the improved sources but also the unimproved sources in to subcategories. With that division, we have a total of 6 categories of water sources: three improved and three unimproved. In Table 1, we show these 6 regrouped categories, as well as the 13 original categories as subcategories of those. Water sources listed from 1 to 3 are the three categories of

improved water sources and those listed from 4 to 6 are the three categories of unimproved water sources. Columns (1), (2) and (3) display the number of children in the sample that use each water source, the respective percentage and the percentage of children in that water source category who had diarrhea during the two weeks prior to the survey.

For the analysis of sanitation facilities, similar to the case of water sources, we first construct a binary indicator by aggregating the 14 mutually exclusive categories of sanitation (toilet) facilities, shown in Table 2, into two groups, “improved” and “unimproved. We then further divide the “unimproved” sanitation facilities into two subcategories (“no private toilet” and “unimproved private toilet”) resulting in a trichotomous variable. Under the “no private toilet” category, we have the households who use a shared facility as well as those who practice open defecation.¹¹ The “unimproved private toilet” category includes various privately-owned unimproved toilet facilities. Unlike the water sources case, here we leave the improved sanitation facility in one category. Given larger percentage of households in Ethiopia defecated openly or shared toilet with other households than used any type of private toilet at the time the data was collected, we find it more sensible to investigate the effect of having a private toilet versus not having than looking at the effect of various types of private toilet facilities available at a household. Columns (1), (2) and (3) of Table 2 display the number of children in the sample that use each type of sanitation facility, the respective percentage and the percentage of children in that sanitation facility category who had diarrhea during the two weeks prior to the survey.

¹¹ These two categories were combined under “no private toilet” category because of the following two reasons. First, places where people in Ethiopia practice open defecation such as fields, forests, and bushes can be loosely considered as shared facilities as those places are often shared by many people. Second, a comparison of shared toilets and open defecation shows that shared toilets could, in fact, result in a worse outcome than open defecation. However, for obvious reasons, no policymaker wants to suggest open defecation for families that share a toilet and both groups have a strong need of a better sanitation facility from a policy perspective. We, therefore, combine these two groups together under “no private toilet” category.

Control Variables

We include the following control variables in the first stage of the of the ATE estimation: an indicator variable to identify rural households, indicators to identify each region in Ethiopia, the number of household members, years of schooling of father, years of schooling of mother and a wealth index. Descriptive statistics of these variables are shown in Table 3. The seven indicator variables to identify region (state) of residence in Ethiopia is expected to control for unobservable factors affecting access to water and sanitation facilities including but not limited to differences in culture, religion, political views, and the level of infrastructural development.¹². First stage estimation results are shown in Appendix A, available online.

The wealth index that we use as a control is not what we find in the original dataset constructed through principal component analysis using several indicators of wealth. The index available in the dataset has been constructed using water sources and sanitation facilities among several other variables, and therefore we cannot use it as an explanatory variable in our models. Considering that, we follow previous researches to construct a new wealth index for this study following a procedure similar to the original, but not using the water source or sanitation related variables. Variables in our wealth index include dummy variables that indicate ownership status of a household of land, bed, house, car, TV, refrigerator, and other valuable properties. Following Novak (2014), we use the natural log of the wealth index as an explanatory variable, based on the assumption that wealth would have diminishing effect in its ability to affect the choice of a household's water source or sanitation facility.

¹²Ethiopia has 11 regional states (See the provisional map of Ethiopia in Chart 1) but we combined the three urban districts enclaved within the Oromia region, Addis Ababa, Harari and Dire Dawa with Oromia, excluded region in our regressions.

Despite the broad list of controls included, there can still be many unobserved factors that could limit causal interpretability of our results. Since we cannot test if any variable determining the treatment and correlated with the outcome is missing in our first stage model, we use indirect methods to check whether the conditions for reliable estimation of treatment effects using this technique are satisfied. The first condition is sufficient balance on observed covariates between the treatment group and the control group. To ensure a good balance, the B statistic, the standardized mean difference of the propensity scores in treated and control groups, should be closer to zero and the R statistic, i.e. the variance ratio of propensity scores of treated and untreated (control) observations, should be closer to 1 (Rubin, 2001). A variance ratio less than or equal to 0.5 or greater than 2 indicates severe covariate imbalance. In addition to the balance of covariates that jointly determine the outcome and the treatment, we also check the balance of some additional variables which determine the outcome but not the treatment. Under randomization, we can expect these additional variables also to be balanced. These variables include age and the sex of the child, whether mother is pregnant, whether mother is breastfeeding and whether a newborn is living in the household. These statistics are presented in Appendix B, which is available online. Overall, both the B and R statistics are in a desired range for credible estimation.

The second condition is sufficient overlap in propensity scores between the treatment and control groups. To check this, we inspect the histograms of propensity scores of the treated and control groups and exclude the observations outside the range of common support before executing our matching algorithm. These histograms are also shown in Appendix B.

4. Results and Discussion

The Effect of Water Sources on Diarrhea

In Table 4, we present the ATEs of different types of water sources on the incidence of diarrhea. The estimate in Panel A is based on the traditional categorization of all water sources into two groups and it shows the ATE of upgrading an unimproved water source to an improved one. We do not find statistically significant evidence that improved water sources would reduce diarrhea prevalence. We hypothesize that the insignificant treatment effect in this case is due to an ‘aggregation effect’ which arises from grouping water sources of different qualities under a single category called ‘improved’. To examine whether this hypothesis is correct, we further categorize the improved water sources into three subcategories: ‘piped in to dwelling/yard/plot’, ‘public tap’ and ‘other improved’ and check whether one or more of these categories show(s) a statistically significant treatment effect.

In Panel B, we present the ATEs of each of these three improved water sources relative to unimproved water sources grouped in a single category. We also present the ATEs amongst the improved source. As the results show, only one of the improved water sources, water piped into dwelling/yard/plot, has a statistically significant ATE relative to unimproved water sources as well as the other two improved sources. In particular, water piped into dwelling/yard/plot is associated with 5.1, 18.2 and 13.0 percentage points less incidence of diarrhea compared to public tap/standpipe, other improved sources and unimproved sources, respectively. This finding supports our argument of heterogeneous effects. It is interesting to note here that the largest ATE of water piped into dwelling/yard/plot is relative to the other improved sources (not against the unimproved sources) indicating that the other improved sources could be inferior even compared to the unimproved sources. Moreover, the ATE of moving to the ‘public tap’ category from ‘other

improved' category is associated with a 4.7 percentage point reduction in diarrhea while the ATE from moving from 'unimproved' category has no statistically significant association suggesting that the so called 'unimproved' sources are not so bad as some of the improved sources. This finding is consistent with Novak (2014) who also finds heterogeneous effects of different types of improved water sources.

The results in panels A and B of Table 4 support our claim that grouping all improved sources together could mask the potential positive impact of some of the improved sources in reducing diarrhea.

In Panel C we present the ATEs when improved and unimproved sources are each subdivided into three categories. We find that water piped into dwelling/yard/plot is the only water source with statistically significant ATEs against all other categories, confirming our finding in Panel B. In particular, water piped into dwelling/yard/plot is associated with 5.2, 13.0, 11.3, 14.4 and 4.6 percentage points less incidence of diarrhea compared to public tap/standpipe, other improved sources, unprotected well, unprotected spring and other unimproved sources, respectively. The heterogeneous effects of water piped into dwelling/yard/plot relative to various unimproved water sources suggests that aggregating all types of unimproved sources in one category could be misleading. It is interesting to find that some of the water sources traditionally considered to be clean could be inferior to some of the unimproved sources in terms of their impact in reducing diarrhea.

Our findings might be useful for donors and governments investigating more effective ways of dealing with the incidence of diarrhea in developing countries. Yet, this result could at least partially be an artifact of unintentional misreporting (Rosenman et al., 2011) since the survey respondents may find it hard to understand the lines of demarcation between unprotected wells

and protected wells or unprotected springs and protected springs leading to misclassification error in data. As a result, some of the wells and springs that the respondents have identified as ‘protected’ and we have grouped within the ‘other improved’ category may in fact be worse than some other wells and springs they have identified as ‘unprotected’ resulting in a misleading result that the ATE of moving from the ‘other unimproved’ category to ‘other improved’ category is favorable as we see in Panel B or insignificant as we see in Panel C.¹³

Overall, the results show that reducing diarrhea in Ethiopia effectively is possible only through provision of water piped into dwellings, yards or plots. We find no evidence that the other water sources, traditionally classified as ‘improved’, play any significant role in reducing diarrhea. Besides the quality of the water sources, other factors such as distance of the water sources from residence, time spent fetching water, availability of clean storage facilities in the household, etc. could explain the ineffectiveness of the other improved water sources in reducing diarrhea. Due to lack of information, we investigate only the role of time spent fetching water and found no qualitative difference between these results (not reported) and our main findings.

Effect of Sanitation Facilities on Diarrhea

For investigating the effect of sanitation facilities on incidence of diarrhea, we follow the same steps as in our analysis of water sources. At first, we categorize sanitation facilities as “improved” and “unimproved” following the standard definition to find no statistically significant ATE as shown in Panel A of Table 5. Having seen no difference between the two traditional categories and suspecting aggregation error as in the case of water sources, we next divide the unimproved category to two sub-categories: no private toilet and unimproved private toilet, increasing the number of categories to three as explained in Section 3. The ATEs from this

¹³ In addition to this, measurement error in data can affect overall reliability of our results too.

trichotomous classification, reported in Panel B of Table 5 are more intuitive with facilities generally considered better resulting in improvements compared to the facilities considered worse. We see a statistically significant 3.0 percentage point reduction in incidence of diarrhea when a household without any private toilet starts to use some private facility even if that facility has traditionally been classified as an unimproved one. However, the gains from an improved toilet are larger (4.3 percentage points). At the same time, the difference between having an unimproved private toilet versus having an improved one is not statistically significant.

Comparing the treatment effects of sanitation facilities and access to water sources on diarrhea, we note two important differences. First, while any private sanitation facility (toilet) helps the objective, not necessarily unimproved one, only one type of improved water sources (water piped into dwelling/yard/plot) is effective in reducing diarrhea incidence. Second, the treatment effects of water sources (4.6 to 14.4 percentages points reduction) are larger in magnitude compared to those of the sanitation facilities (3.0 to 4.5 percentage points reduction).¹⁴

The next policy-relevant question we attempt to address is whether it is more effective if the same household is twice-upgraded with a toilet facility and piped water than upgrading two different households. We test this in two ways: 1) we estimate the treatment effects of having a toilet with and without being complemented with piped water and investigate the differences in the ATEs; and 2) we separately estimate the treatment effects of having piped water for households having no private toilet, unimproved private toilet and improved toilet and investigate the differences in the ATEs. These results are shown in Table 6.

¹⁴ Capuno et al. (2015) find that an improvement in sanitation facility has a larger effect than the effect of upgrading the water source. The treatment in their case, however, is whether households were provided with their own flush toilets, which is not the case in our paper. Begum et al. (2011) find no significant effect of either an improved water source or improved sanitation facility alone. However, they find a significant reduction in diarrhea when a household has both improved water and improved sanitation. Roushdi and Sieverding (2016) only investigate the effect of an improved water source.

As the results presented in Panel A of Table 6 suggest, having a toilet has a significant treatment effect (2.8 percentage point reduction in diarrhea) for households with water piped into their dwelling, yard or plot. The treatment effect is a marginally insignificant 2.2 percentage point reduction for households who do not have water piped into their dwelling, yard or plot. This result suggests the presence of some complementary effects. However, we see a contrasting result when we look at the effects of having water piped into dwelling, yard or plot for households at different levels of sanitation facilities. As shown in Panel B, those with an improved sanitation facility gain less from water piped to the dwelling, yard or plot than those without an improved sanitation facility. The respective ATEs are 5.7, 14.1 and 12.8 percentage point reductions in incidence of diarrhea for those with an improved toilet, with an unimproved private toilet and with no private toilet, respectively. This very last result, however, should be interpreted with caution since we see limited overlap across groups (See Appendix B).

All our baseline results are robust to application of a tight caliper. We estimate each ATE that we present and discuss here using appropriate calipers based on the criteria that we explain at the end of Section 2 and those results are presented in Tables 7, 8 and 9.

5. Summary and Conclusions

We investigate the impacts of the drinking water sources and sanitation facilities on child health in Ethiopia using propensity score matching allowing multiple treatments. The results show heterogeneous impacts of both improved and unimproved water sources on child health, suggesting that the traditional way of categorizing water sources and sanitation facilities into a binary variable as “improved” and “unimproved” could be misleading. While water piped into dwelling, yard or plot leads to a large reduction in diarrhea incidence relative to all improved and unimproved sources, other types of water sources, generally considered as clean, are actually not

different from the unimproved sources in terms of reducing diarrhea. Meanwhile, having any toilet, even if it considered unimproved, has significant effects in reducing incidence of diarrhea. Finally, we have mixed evidence about the comparative effectiveness of policy interventions to provide either a toilet or piped water to two different households and providing both facilities to the same household.

We haven't, however, investigated the cost effectiveness of various potential policy interventions here. Though the gains from piped water appear to be larger than the gains from having a toilet, in general, infrastructure for piped water needs larger financial investments compared to the cost of a toilet and therefore may not outperform the latter in terms of its cost-effectiveness. We leave this for future research.

One of the caveats of this type of studies which use survey data is the potential measurement error. The respondent may not list her water source and/or the sanitation facility under the correct category if the definitional differences are not clear to her. Another potential cause of misclassification error could arise from the fact that the survey only collects information about the primary source of drinking water. Some households, particularly those who travel a longer distance to fetch water, are likely to use water from additional sources frequently and the driver of diarrhea in such households could be one of these secondary water sources. All those forms of measurement error could bias our results but the direction of such bias is not clear.

ACKNOWLEDGMENTS

We thank the editor Andy McKay and two anonymous referees for their detailed comments and many great suggestions which helped to improve this manuscript substantially. We also thank the seminar participants at the 87th Annual Meeting of the Indiana Academy of the Social Sciences

held in Westville, and Annual Meeting of the Midwest Economic Association held in Cincinnati in 2017.

References

- Begum, S., Ahmed, M., & Sen, B. (2011). Do water and sanitation interventions reduce childhood diarrhoea? New evidence from Bangladesh. *The Bangladesh Development Studies*, 1-30.
- Capuno, J. J., Tan Jr, C. A. R., & Fabella, V. M. (2015). Do piped water and flush toilets prevent child diarrhea in rural Philippines? *Asia Pacific Journal of Public Health*, 27(2), NP2122-NP2132.
- Centers of Disease Control and Prevention (2016) https://www.cdc.gov/healthywater/global/wash_statistics.html#four last accessed on 09/08/2016.
- Cochran, W. G., and Rubin, D. B. (1973). Controlling Bias in Observational Studies: A Review. *Sankhyā: The Indian Journal of Statistics, Series A* (1961-2002), 35, 417-446.
- Fan V, Mahal A. 2011. What prevents child diarrhoea? The impacts of water supply, toilets, and hand-washing in rural India. *Journal of Development Effectiveness* 3: 340–370.
- Stuart, E. A., & Rubin, D. B. (2004). Matching methods for causal inference: Designing observational studies. *Harvard University Department of Statistics* mimeo.
- Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164, 73-84.
- Jalan, J., & Ravallion, M. (2003). Does piped water reduce diarrhea for children in rural India?. *Journal of econometrics*, 112(1), 153-173.
- Kumar, S., & Vollmer, S. (2013). Does access to improved sanitation reduce childhood diarrhea in rural India?. *Health Economics*, 22(4), 410-427.
- Lechner, M. (2002). Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *The review of economics and statistics*, 84(2), 205-220.
- Novak, L. (2014). The impact of access to water on child health in Senegal. *Review of Development Economics*, 18(3), 431-444.

- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural & Healthcare Research*, 2(4), 320–332.
<http://doi.org/10.1504/IJBHR.2011.043414>
- Roushdy, R., & Sieverding, M. (2016). Improved water and child health in Egypt: impact of interrupted water supply and storage of household water on the prevalence of diarrhea. *EMHJ-Eastern Mediterranean Health Journal*, 22(1), 5-19.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688-701.
DOI: <http://dx.doi.org/10.1037/h0037350>.
- Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, 31(2), 161-170.
- UNICEF (2015). Monitoring the Situation of Children and Women, Committing to Child Survival: A Promise Renewed. *Progress Report*.
- World Health Organization, & UNICEF. (2017). Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines.
<http://www.who.int/mediacentre/news/releases/2017/launch-version-report-jmp-water-sanitation-hygiene.pdf>

Table 1: Incidence of Diarrhea and the Water Source Used

Source of drinking water	Number of observations	Percentage of observations	Percentage that had diarrhea
1. <i>Piped to dwelling/yard/plot total</i>	593	6.23	10.8
Piped into dwelling	51	0.56	13.46
Piped into yard/plot	542	5.97	10.55
2. <i>Public tap/standpipe</i>	1667	18.74	13.38
3. <i>Other improved total</i>	2183	24.44	16.56
Protected well	915	10.43	15.09
Tube well or borehole	528	5.88	20.11
Protected spring	666	7.26	16.29
Rain water	74	0.87	12.50
4. <i>Unprotected well</i>	540	6.02	18.56
5. <i>Unprotected spring</i>	1903	20.92	15.61
6. <i>Other unprotected total</i>	2094	23.35	15.57
Surface water	1945	21.58	15.64
Cart with small tank/drum	82	1.01	15.05
Tanker-truck	65	0.74	19.12
Bottled water	2	0.02	0.00
Observations	8,980	100.00	

Table 2: Incidence of Diarrhea and the Type of Sanitation Facility

Type of Sanitation Facility	Number of observations	Percentage of observations	Percentage that had diarrhea
1. <i>Improved toilet</i>	1,163	8.24	11.73
Flush to piped sewer system	43	0.47	6.98
Flush to septic tank	34	0.37	14.71
Flush to pit latrine	78	0.85	11.54
Flush, don't know where	7	0.08	0.00
Ventilated improved pit latrine	98	1.06	15.31
Pit latrine with slab	278	3.02	7.20
Composting toilet	221	2.40	16.74
2. <i>Unimproved private toilet</i>	3,304	24.14	14.25
Pit latrine without slab/open pit	2,213	24.01	14.28
Flush to somewhere else	4	0.04	0.00
Bucket toilet	2	0.02	0.00
Hanging toilet/latrine	1	0.01	0.00
Other	5	0.05	20.00
3. <i>No private toilet</i>	6,232	67.62	16.13
No facility/bush/field	4,749	51.53	15.96
Shared facility	1483	16.09	16.66
Observations	9,216	100.00	

Note: A shared facility can be any of the improved or unimproved types.

Table 3: Summary Statistics

Variable	Mean	Std. Dev.
Diarrhea Prevalence	0.151	0.358
Rural	0.844	0.362
Region		
Tigray	0.102	0.302
Afar	0.098	0.298
Amhara	0.11	0.313
Oromia	0.159	0.366
Somali	0.08	0.272
Benishangul	0.088	0.284
SNNP*	0.145	0.353
Gambella	0.063	0.243
Harari	0.057	0.233
Addis Ababa	0.034	0.181
Dire Dawa	0.059	0.237
Members of household	6.146	2.229
Log of Wealth Index	1.073	0.404
Mother's years of schooling	1.451	3.045
Father's years of schooling	2.808	3.989
Observations	8, 980	

* Southern Nations, Nationalities, and Peoples' Region.

Chart 1: Regions (States) of Ethiopia



Table 4: The Average Treatment Effects of Moving from One Type of Water Source to Another (Nearest neighbor matching with replacement)

Panel A: Two categories of water sources (improved vs unimproved)					
Water source	Unimproved				
Improved	-0.002 (0.012) [8980/8980]				
Panel B: Four categories of water sources (improved sources sub-divided into three)					
Water source	Unimproved sources	Other improved sources	Public tap (standpipe)		
Other improved sources	0.007 (0.012) [6048/6720]				
Public tap/standpipe	-0.007 -0.016 [6144/6204]	-0.047* (0.027) [3780/3850]			
Piped into yard/plot/dwelling	-.130*** (0.018) [5064/5130]	-.182*** (0.039) [2776/2776]	-.051*** (0.018) [2187/2260]		
Panel C: Six categories of water sources (3 improved and 3 unimproved)					
Water source	Other unimproved sources	Unprotected spring	Unprotected well	Other improved sources	Public tap (standpipe)
Other improved sources	0.028* (0.015) [4269/4277]	-0.008 (0.020) [3682/4086]	-0.003 (0.026) [1678/2723]		
Public tap/standpipe	0.007 (0.017) [3360/3761]	-0.018 (0.018) [2903/3570]	-0.033 (0.030) [1297/2207]	-0.027 (0.024) [3657/3850]	
Piped into dwelling/yard/plot	-0.046*** (0.017) [1395/2687]	-0.144*** (0.018) [1943/2496]	-0.113*** (0.018) [651/1133]	-0.130*** (0.016) [2273/2776]	-0.052*** (0.018) [2177/2260]

Notes: Standard errors are in parentheses. Number of observations before and after matching is inside square brackets. Average treatment effects are for moving from the water source given by the column to the water source given by the row. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Presents treatment effect using the traditional categorization of the water sources.

Panel B: Presents treatment effects with three categories of improved sources and all unimproved sources grouped together

Panel C: Presents treatment effects with six categories of water sources (both improved and unimproved sources are sub-divided).

Table 5: The Average Treatment Effects of Moving from One Type of Sanitation Facility to Another (Nearest neighbor matching with replacement)

Panel A: Two categories of sanitation facilities (improved vs. unimproved)

Sanitation facility	Unimproved sanitation
Improved sanitation	0.004 (0.022) [9198/9216]

Panel B: Three categories of sanitation facilities (unimproved facilities sub-divided into two)

Sanitation facility	No private toilet	Unimproved private toilet
Unimproved private toilet	-0.029* (0.017) [8167/8457]	
Improved private toilet	-0.043** (0.023) [6845/6991]	-0.024 (0.017) [2891/2984]

Notes: Standard errors are in parentheses. Number of observations before and after matching is inside square brackets. Average treatment effects are for moving from the water source given by the column to the water source given by the row. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: The Average Treatment Effects of Complementarities (Nearest neighbor matching with replacement)

Panel A: Two categories of sanitation facilities (improved vs. unimproved)			
Treatment	Water piped into dwelling/yard/plot	Water not piped into dwelling/yard/plot	
Having private toilet vs. not	-0.028* (0.0316) [596/602]	-0.022 (0.030) [8574/8614]	
Panel B: Three categories of sanitation facilities (unimproved facilities subdivided into two)			
Treatment	No private toilet sources	Unimproved private toilet	Improved toilet
Water piped into dwelling/yard/plot vs. not	-0.128 (0.013) [5194/6991]	-0.141 (0.009) [2205/2225]	-0.057*** (0.028) [722/759]

Notes: Standard errors are in parentheses. Number of observations before and after matching is inside square brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Heterogeneous average treatment effects of having a private toilet compared to those having none for the households with and without water piped to dwelling, yard or plot. The row represents treatment versus control (the treatment dummy variable takes the value 1 if the household has private toilet and 0 otherwise). The two columns represent the samples over which the estimation is conducted.

Panel B: Heterogeneous average treatment effects of having water piped to dwelling, yard or plot for the households with no private toilet, unimproved private toilet and improved toilet. The row represents treatment versus control (the treatment dummy variable takes the value 1 if the household has water piped into dwelling/yard/plot and 0 otherwise). The three columns represent the samples over which the estimation is conducted.

Table 7: The Average Treatment Effects of Moving from One Type of Water Source to Another (Nearest neighbor matching with replacement and the caliper is set at $\alpha = 0.2$)

Panel A: Two categories of water sources (improved vs unimproved)					
Water source	Unimproved				
Improved	-0.002 (0.012) [8979/8980]				
Panel B: Four categories of water sources (improved sources sub-divided into three)					
Water source	Unimproved sources	Other improved sources	Public tap (standpipe)		
Other improved sources	0.006 (0.012) [6010/6720]				
Public tap/standpipe	-0.007 -0.016 [6104/6204]	-0.047* (0.027) [3761/3850]			
Piped into yard/plot/dwelling	-.138*** (0.014) [4848/5130]	-.182*** (0.039) [2723/2776]	-.051*** (0.018) [2185/2260]		
Panel C: Six categories of water sources (3 improved and 3 unimproved)					
Water source	Other unimproved sources	Unprotected spring	Unprotected well	Other improved sources	Public tap (standpipe)
Other improved sources	0.027* (0.015) [4232/4277]	-0.012 (0.018) [3622/4086]	-0.002 (0.027) [1586/2723]		
Public tap/standpipe	0.005 (0.018) [3278/3761]	-0.019 (0.017) [2854/3570]	-0.033 (0.030) [1297/2207]	-0.027 (0.024) [3657/3850]	
Piped into dwelling/yard/plot	-0.062** (0.013) [1216/2687]	-0.146*** (0.01) [2100/2496]	-0.148*** (0.018) [571/1133]	-0.151*** (0.012) [2134/2776]	-0.052*** (0.018) [2175/2260]

Notes: Standard errors are in parentheses. Number of observations before and after matching is inside square brackets. Average treatment effects are for moving from the water source given by the column to the water source given by the row. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Presents treatment effect using the traditional categorization of the water sources.

Panel B: Presents treatment effects with three categories of improved sources and all unimproved sources grouped together

Panel C: Presents results treatment effects with six categories of water sources (both improved and unimproved sources are sub-divided).

Table 8: The Average Treatment Effects of Moving from One Type of Sanitation Facility to Another (Nearest neighbor matching with replacement and the caliper is set at $\alpha = 0.2$)

Panel A: Two categories of sanitation facilities (improved vs unimproved)

Sanitation facility	Unimproved sanitation
Improved sanitation	0.015 (0.033) [501/9216]

Panel B: Three categories of sanitation facilities (unimproved facilities sub-divided into two)

Sanitation facility	No private toilet sources	Unimproved private toilet
Unimproved toilet	-0.030* (0.017) [8232/8457]	
Improved toilet	-0.045** (0.023) [6845/6991]	-0.023 (0.019) [2878/2984]

Notes: Standard errors are in parentheses. Number of observations before and after matching are inside square brackets. Average treatment effects are for moving from the water source given by the column to the water source given by the row. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Presents treatment effect using the traditional categorization of sanitation facilities.

Panel B: Presents treatment effects with three categories of sanitation facilities

Table 9: The Average Treatment Effects of Complementarities (Nearest neighbor matching with replacement and the caliper is set at $\alpha = 0.2$)

Panel A: Two categories of sanitation facilities(improved vs unimproved)

Treatment	Water Piped into dwelling/yard/plot	Water not piped into dwelling/yard/plot	Both groups
Having private toilet vs not	-0.035 (0.03) [546/602]	0.026 (0.016) [8509/8614]	0.035 (0.013) [9117/9216]

Panel B: Three categories of sanitation facilities (unimproved facilities subdivided into two)

Treatment	No private toilet sources	Unimproved private toilet	Improved toilet	All three groups
Water piped into dwelling/yard/plot vs not	-0.139*** (0.011) [6232/6991]	-0.139 (0.403) [2022/2225]	-0.087*** (0.019) [632/759]	-0.135*** (0.009) [7945/9216]

Notes: Standard errors are in parentheses. Number of observations before and after matching are inside square brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Heterogeneous average treatment effects of having a private toilet compared to having none for the households with and without water piped to dwelling, yard or plot. The row represents treatment versus control (the treatment dummy variable takes the value 1 if the household has private toilet and 0 otherwise). The two columns represent the samples over which the estimation is conducted.

Panel B: Heterogeneous average treatment effects of having water piped to dwelling, yard or plot a private toilet vs. not having for the households with no private toilet, unimproved private toilet and improved toilet. The row represents treatment versus control (the treatment dummy variable takes the value 1 if the household has private toilet and 0 otherwise). The two columns represent the samples over which the estimation is conducted.